A leap into the chemical space of protein-protein interaction inhibitors.

Bruno O. Villoutreix, Céline Labbé, David Lagorce, Guillaume Laconde, Olivier Sperandio

Title: ”A leap into the chemical space of Protein-Protein Interaction inhibitors.”

Villoutreix B.O[1,2]., Labbé C.[1,2], David Lagorce[1,2], Laconde G.[1,2], Sperandio O[1,2 *].
[1] INSERM, U973, F-75205 Paris, France
[2] Univ Paris Diderot, Sorbonne Paris Cité, UMRS 973, MTi, F-75205 Paris, France
* corresponding author: olivier.sperandio@inserm.fr

Abstract

Protein-protein interactions (PPI) are involved in vital cellular processes and are therefore associated to a growing number of diseases. But working with them as therapeutic targets comes with some major hurdles that require substantial mutations from our way to design to drugs on historical targets such as enzymes and G-Protein Coupled Receptor (GPCR). Among the numerous ways we could improve our methodologies to maximize the potential of developing new chemical entities on PPI targets, is the fundamental question of what type of compounds should we use to identify the first hits and among which chemical space should we navigate to optimize them to the drug candidate stage. In this review article, we cover different aspects on PPI but with the aim to gain some insights into the specific nature of the chemical space of PPI inhibitors. We describe the work of different groups to highlight such properties and discuss their respective approach. We finally discuss a case study in which we describe the properties of a set of 115 PPI inhibitors that we compare to a reference set of 1730 enzyme inhibitors. This case study highlights interesting properties such as the unfortunate price that still needs to be paid by PPI inhibitors in terms of molecular weight, hydrophobicity, and aromaticity in order to reach a critical level of activity. But it also shows that not all PPI targets are equivalent, and that some PPI targets can demonstrate a better druggability by illustrating the better drug likeness of their associated inhibitors.

Abbreviation

PPI	Protein-Protein Interaction
iPPI	small molecule inhibitor of Protein-Protein Interaction
RO5	Lipinski’s Rule of Five
ADME/Tox	Absorption Distribution Metabolism Excretion/ Toxicity
LE	Ligand Efficiency: Gibbs Free Energy of binding/(number of heavy atoms)
LEHA	Ligand Efficiency: pIC50/(number of heavy atoms) LEHA = pIC50/(number of heavy atom)
LLE	Lipophilic Efficiency LLE = pIC50 - AlogP
ΔASA	Solvent accessible surface area

Introduction

As the pharmaceutical industry and the academic system are under a dramatic pressure to offer more efficient and safer drugs to the patients on a growing number of diseases, there are urgent needs to boost pharmaceutical innovations (novelty of effectiveness), and developing new conceptual frameworks/methodological tools to assist the discovery of novel drugs. This may come, for instance, with the study of new

therapeutics targets. But it requires sufficient knowledge prior to embark to such expensive investigations. To this end, tremendous efforts in the clarification of the role of Protein-Protein Interactions (PPI) in the health and disease states are currently taking place. Indeed, PPI, by their implication in numerous cellular mechanisms are more and more linked to a growing number of diseases and therefore represent a remarkable pool of putative therapeutic targets. With an estimation ranging from 130,000[1] to about 650,000[2] PPI in human, excluding trans-organism PPI (mainly important for infectious diseases), the number of potential targets is exceptional, although not all those PPI might represent a druggable target capable of binding a small molecular drug nor some might be relevant to pursue given their role in a particular biological pathway. Nevertheless, despite their potential at the biological level, PPI have been thought for many years to be undruggable[3] or too risky using low molecular weight molecules. The first reason involved the topological nature of their interface, assumed to be too large and/or too hydrophobic to be compatible with the physicochemical properties of most known small molecular drug candidates (these properties involve mainly molecular weight and octanol water partition coefficient most often reported as logP). The second reason is the nature of the chemical collections that are used presently in high throughput screening campaigns that were essentially designed for regular targets such as G-protein Coupled Receptors (GPCR), ion channels and enzymes. Yet, the growing number of success stories in the modulation of PPI with small molecules including peptido-mimetic and "regular" chemical compounds, allows now the scientific community to start gaining insight into the general properties that an inhibitor of PPI **(iPPI)** should possess, at least in the light of existing successful examples. This indeed represents valuable information about their associated chemical space and potential strategies to explore further this challenging class of novel therapeutic targets.

In this review, we will describe our present knowledge of the iPPI chemical space and how we could derive in the near future new ways of both filtering and prioritizing chemical synthesis of putative iPPI with acceptable pharmacokinetic profiles. But before going further we should start by clarifying and explaining some keys concepts on PPI and the type of pharmacological modulation considered will be summarized in order to properly assess the corresponding physico-chemical profile of iPPI.

Types of interaction

PPI are highly heterogeneous and can be found in many forms. Indeed, depending whether they are obligate or non-obligate, permanent or transient, involved in weak or strong interactions, different topological, kinetic (k_{on}, k_{off}) and energetic profiles can be anticipated and therefore can have various consequences on the physico-chemical and pharmacological profiles of iPPI. The concentration in cells of both protein partners (or more if multimers) may also become prevalent and have different consequences on the pharmacological modulation one might want to achieve.

One can also argue on the type of protein partners involved. Indeed, it is not unusual to consider as a PPI an interaction between a protease and its substrate or with some protein inhibitor (eg eglin C with different serine proteases). The debate could therefore be whether one should consider the catalytic site of a protease as the location of a true transient interaction with its substrate or as a site of chemical reaction intermediate (here tetrahedral intermediate) in which for a very short duration the two proteins are chemically bound to each other with evidently very different consequences in terms of binding profiles. In addition, there might be confusions about the meaning of protein-

protein interactions. The type of interactions that we address here are true interaction in a sense that the interface zone involved does not overlap with a catalytic site or with pocket known to bind a small compound. For instance, some GPCRs can bind a number of external and endogenous compounds in some well known binding cavities, some of these binders are small ligands/substrates, others are peptides or else entire proteins but these binding pockets have indeed evolved to bind large and small ligands but do not fit into the type of interactions that we are investigating here. Again, if we consider BPTI, this protein binds and blocks the catalytic site of a serine protease, compounds that will block this interaction by interacting with the serine protease active site can not be considered a true PPI inhibitors as this pocket has also evolved to bind small ligands and as such there will be little to learn that we already know in term of chemistry and chemical space specificity.

Properties of the interface

Various studies have described the topological properties of PPI at the interface[4, 5], and the purpose of this review is not to go through all of them but rather to cite some and attempt to extract what could be appropriate in terms of anticipated physico-chemical properties for an ideal iPPI. In an influential study, Jones and Thornton by analyzing a set of 59 complexes found that the solvent accessible surface area (ΔASA) spans from 368 $Å^2$ to 4746 $Å^2$ in homodimers and from 639 $Å^2$ to 3228 $Å^2$ in heterocomplexes[6]. Yet, ten of the PPI within that study were based on enzyme/inhibitor complexes, with an average ΔASA of 785 $Å^2$ among the smallest of the full dataset. Nevertheless, this gives a conceptual idea of the size an iPPI might need to reach to mimic the presence of the native partner to prevent its interaction. Fortunately, the seminal work of Wells[7] using alanine scanning has highlighted the presence of key residues at the interface on which resides most of the binding energy. This confirmed that only part of the interaction patch really contributes to the partners' binding such that, as long as the iPPI is capable of disrupting those hot spots, the modulation is at least theoretically achievable. The growing number of success stories, and especially the work achieved by the fragment-based community has supported this observation on various cases.

The contributions of various residues at the interface can also provide valuable information that could be translated into chemical properties of iPPI. Several studies have attempted to assess the over-representation of key amino acids within protein interfaces. These studies have for example highlighted the importance of Trp/Met/Phe clusters within hotspots regardless of the type of PPI considered[8] or more specifically the presence of Phe, Trp, Leu and Arg in α-helix mediated interaction[9]. Other studies[10] have analyzed aromatic-aromatic interactions between protein partners and described geometrical distribution and orientation of the aromatic rings in place. Finally, some studies have attempted to overlap the protein binding sites with small molecules within families of structures including PPI[11] and others start to propose methods and tools to predict what interface could be adapted for low molecular weight compound modulation[12]. The combination of this information could be used in some cases to derive physico-chemical properties matching both those of the corresponding residues and more generally those of the associated binding pockets.

Orthosteric versus allosteric modulation

Besides the interface and type of interaction, the type of modulation is also to be considered. Inhibitors can be for instance direct/orthosteric inhibitors binding essentially at the interface between two proteins (very few small organic compounds

orally available are known at this time) or can be allosteric inhibitors, thereby binding at, near or far away from the interface in cavities with recognition characteristics that could be highly similar to enzyme active sites. It would seem logical to explore and attempt to rationalize the properties of the direct protein-protein inhibitors (they also may act via small structural changes) because the chemistry of these ones is likely to be novel as the compounds have to bind to surfaces that have not evolved to bind small molecules. Also allosteric compounds have been used for many years but the mechanism of action has often been found retrospectively because of the complexity of the mechanisms involved. Thus, although allosteric inhibitors are interesting from an energetic standpoint, for their selectivity…etc, this is not the type of compounds that we will be dealing with in this review to explore novel regions of the chemical space (i.e., these regions might be already known or the complexity of the molecular events taking place impede, at our present level of knowledge, rationalization). For instance, maraviroc (a nonpeptidic, small molecule human immunodeficiency virus type 1 (HIV-1) entry inhibitor) that binds to the CC chemokine receptor 5 (CCR5), a G-protein coupled receptor, prevents the binding of the chemokine CCL3 and the viral envelope glycoprotein gp120 by an allosteric mechanism would not be considered as a true iPPI in our case study as it inhibits the interaction by an allosteric mecamism.

Leap into the iPPI chemical space

Seminal works

Historically, a vast number of studies have been successfully used to rationally design iPPI mostly around privileged structures including peptido-mimetic compounds or specific scaffolds[13-22]. Here we will give a global overview of the various profiling that have been carried out on existing iPPI and present a synthetic depiction of the generic physico-chemical profile of these small molecules.

Besides the legitimate wish to derive physico-chemical properties for iPPI from the topological properties of protein interfaces, a major leap can be undertaken into the chemical space of iPPI by analyzing the successful examples of iPPI themselves. Various initiatives have been undertaken to establish some general trends on the chemical space of iPPI, both to describe its characteristics and to underpin its paradigm shift with respect to those of commercial chemical libraries.

By analyzing 19 iPPI from the literature active on 12 different PPI targets using a principal component analysis with three molecular descriptors (LogP, Molecular Weight and Polar Surface Area), Pagliaro *et al* showed that their chemical space (at least described by those descriptors) did not overlap those of three chemical vendors. Indeed, the inhibitors of only 4 targets out of 12 (Bcl-xL/Bak, MDM2/p53, NGF/p75, and ICAM/LFA) were covered by those commercial collections. They also showed that only 8 out of 19 iPPI could survive Lipinski's RO5 (see below).

In an other study Wells and McClendon have estimated the ligand efficiency of a dozen of iPPI to be around 0.24 kcal/mol compared to those of protease inhibitors (0.25-0.35 kcal/mol) or kinase inhibitors (0.30-0.40 kcal/mol). This would bring the molecular weight of an iPPI having a 10nM K_D to 645 Da, therefore representing a first violation to the Lipinski's RO5.

Several studies have highlighted general and qualitative trends for iPPI. Particularly, reviews from Berg, Fry, and Wilson described iPPI with higher molecular, higher hydrophobicity, higher number of rings, higher aromaticity[13-17].

The road to a rationalization of the iPPI chemical space

With the final goal of improving the poor hit rates usually obtained during the HTS campaigns, several studies have been performed either to propose collections of crystallized iPPI or statistical models able to partially rationalize the global chemical space of iPPI. These endeavors should indeed assist in the selection or synthesis and prioritization of compounds to be screened and thus contribute to enrich the compound libraries in putative protein-protein interface binders.

Concerning the former type of approaches, two different groups have gathered into online databases co-crystallized iPPI. First of this kind, the Timbal database[18] contains 89 iPPI with pharmacological data on 19 PPI targets. Among those, 41 iPPI are co-crystallized, which represents a cumulated number of 10 PPI targets. Similarly, the 2P2I database[19] collects 42 crystallized iPPI with data on both the apo- and the holo-structures on 10 different PPI targets.

Beyond providing to the PPI community collections of iPPI taken from the literature, other initiatives have attempted to propose some rationalization of the iPPI chemical space. The work of Neugebauer[20] used machine-learning techniques on a small set of 25 iPPI and 1135 FDA-approved drugs, as non-iPPI to try to characterize a physico-chemical profile specific of iPPI. They showed in the context of their dataset, that the combination within a decision tree of very few descriptors including SHP2 (known as a shape descriptors introduced by Randic), and the number of ester functions was capable of producing a focused collection with several fold enrichments of inhibitors of protein-protein interactions. Morelli et al. also using some statistical techniques on 39 iPPI from the 2P2I database data have recently proposed a set of thresholds for some commonly used physico-chemical properties in drug development[21]. Their data are characterized by averages values for the molecular weight, AlogP, number of rings and number of Hydrogen bond acceptors of 547 ± 154 Da, 3.99 ± 2.37, 4.44 ± 1.02, 6.62 ± 2.60, respectively.

Recently, our group used a complementary approach[22, 23] to further characterize the physico-chemical profile of iPPI. A set of 66 iPPI “ADME/Tox compatible” (at least after some in silico predictions), and diverse chemically, was fed to a statistical analysis along with a negative dataset of 557 regular drugs that resulted from the same filtering and diversity protocol. The statistical analysis was designed to highlight discriminative descriptors capable of separating the global iPPI population from the regular drug population. Common descriptors such as MW, AlogP, TPSA, nb of rings, etc, were also inspected to collectively provide a full evaluation of the physico-chemical profile of iPPI and therefore gave the first quantified characterization of their chemical space. The main results emphasized an average MW of 421 Da for iPPI versus 341 Da for regular drugs, and an average AlogP of 3.58 versus 2.61. Those results also showed that the number of rings (=4), benzene-like rings (=2), and aromatic bonds (=16) are significantly higher in iPPI than in regular drugs. Finally and importantly, the statistical model, PPIHitProfiler, that was constructed within those two studies, namely a two-descriptor decision tree (using RDF070m, and unsaturation index) demonstrated that the combination of a specific molecular shape and a critical number of 17 multiple bonds is determinant to maximize the iPPI potential of a compound. The model was further validated on experimental screening data taken from PubChem BioAssay on a cumulated number of 500 000 tested compounds across 11 assays. The conclusions of this work were further used to prepare a chemical collection in a recent study that led to the identification of small compound that inhibit the von Willebrand Factor (VWF) A1 - glycoprotein (GP) Ibα interaction[24].

Case Study

In order to propose an update on the major physico-chemical properties of iPPI, we have assembled a chemically diverse set of 115 iPPI that are active on 4 of the most widely studied PPI in the literature, namely the p53/MDM2, Xiap/smac, Bcl-2 family/Bak, and ICAM-1/LFA-1 interactions **(Figure 1, left panel)**. As a reference dataset, we have also considered a representative pool of 1730 chemically diverse inhibitors active on the most widely studied enzymes including some kinases, all taken from the binding database[25] (www.bindingDB.org) along with their binding affinities **(Figure 1, right panel)**. Compounds from the BindingDB are a mix of hits, optimized compounds (lead), and some times drugs depending on the study and the degree of validation of the target. In this reference dataset, the 1730 compounds were taken from an initial set of 46000 active compounds which was submitted to fingerprint-based sampling (ward clustering) to ensure chemical diversity. A series of 24 interpretable molecular descriptors **(Table 1)** such as molecular weight, AlogP, number of sp3 carbon, number of rings, etc… was calculated on the two datasets for estimating the relative characteristics of their respective chemical space and the major properties associated with it. JChem Base was used to generate de novo conformations of the compounds and calculate the descriptors, JChem 5.7, 2011, ChemAxon (www.chemaxon.com).

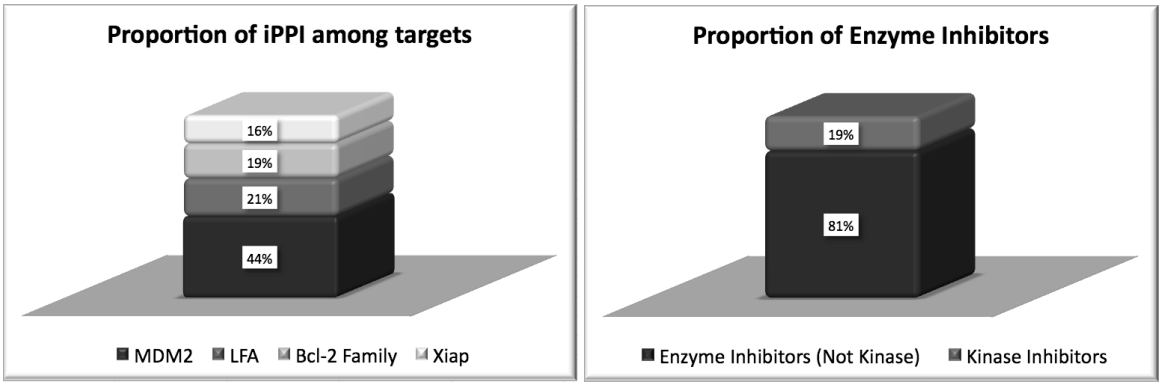


Figure 1: Distribution of iPPI and enzyme inhibitors across their corresponding targets

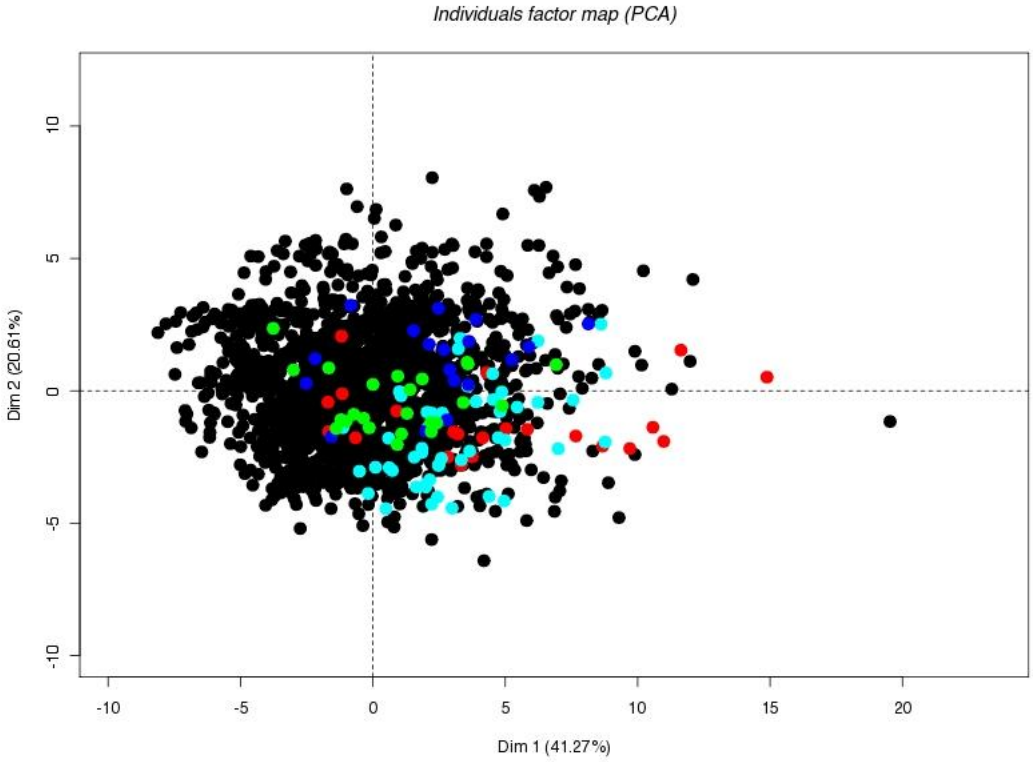
Table 1

N°	Descriptor Code	Descriptor Name	N°	Descriptor Code	Descriptor Name
1	MW	Molecular Weight	13	nCsp3	Nb of sp3 Carbon
2	nAT	Nb of atoms	14	nCIC	Nb of rings
3	nSK	Nb of heavy atoms	15	nBnz	Nb of benzene-like rings
4	nBT	Nb of bonds	16	ARR	Aromatic ratio
5	RBN	Nb of rotatable bonds	17	RDF070m	Radial Distribution Function at 7 Angstroms weighted by atomic mass
6	RBF	Fraction of rotatable bonds	18	nCar	Nb of aromatic carbon
7	nDB	Nb of double bonds	19	nHDon	Nb of H bond donor
8	nAB	Nb of aromatic bonds	20	nHAcc	Nb of H-bond acceptor
9	nC	Nb of Carbon	21	Ui	Unsaturation index
10	nN	Nb of Nitrogen	22	AMR	Molecular refractivity
11	nO	Nb of Oxygen	23	TPSA.Tot.	Topological surface area
12	nX	Nb of Halogen	24	ALOGP	logP

Visualizing the iPPI chemical space

To visualize the chemical space corresponding to those two datasets, we ran a principal component analysis (PCA) using the 24 molecular descriptors listed above on the merged datasets (iPPI + enzyme inhibitors) **(Figure 2)**. Data were scaled to unit variance. The respective positions in chemical space of iPPI (colors dots) on one hand and of the enzyme inhibitors (black dots) on the other hand, within the individual map (top panel), show that there is a slight shift between the two population distribution, the iPPI population being out centered toward the bottom right side of the map. An examination of the variable map (bottom panel) and the descriptors associated to the two first axis of the PCA, that represent more than 60% of the total variance, permit to identify the main physico-chemical properties' discrepancies responsible for this shift. Indeed, the first axis is positively correlated to the size of the compounds (e.g molecular weight, number of heavy atoms, number of carbon, etc). The second axis is clearly negatively correlated to the aromatic ratio i.e to the proportion of aromatic atoms in the compounds. This collectively means that iPPI tend to be heavier and more aromatic than enzyme inhibitors, and that this seems to be the most prevalent characteristics in terms of global variance. But it can also be observed, using the individual map, that not all iPPI are characterized equally by this tendency, depending on the compound but also on the PPI target one considers. For example, MDM2 compounds (cyan) seem to be among the most aromatic iPPI (bottom of the map) while Xiap compounds (blue) are not as aromatic (upper in the map). Also, one can easily see that Bcl-2 inhibitors have very different positions along the first axis and that some of the Bcl-2 compounds (red) are

shifted right because of their significantly higher size, while some of the LFA compounds (green) are quite similar in size with enzyme inhibitors.
The third axis of the PCA, when combined with the two first, represents 73% of the cumulated variance and mostly correlates with the number of sp3 carbon and TPSA. Interestingly, the difference of mean values for those two descriptors was not significant between iPPI and enzyme inhibitors.



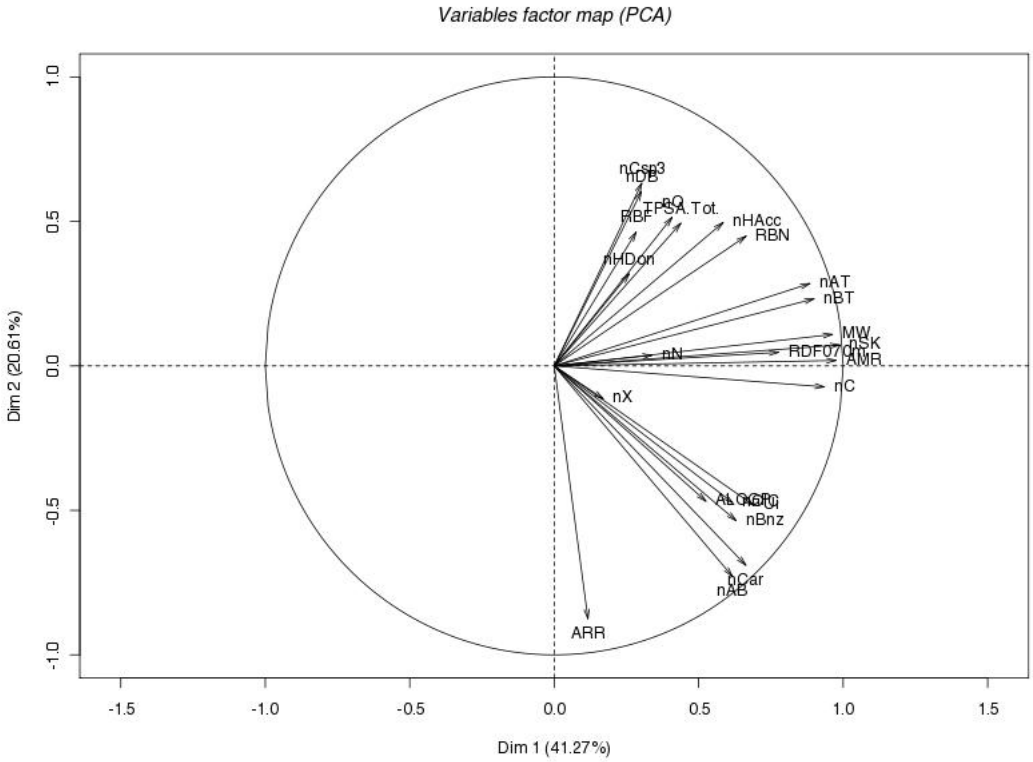


Figure 2: Principal Component Analysis. Individual map (top panel) with enzyme inhibitor in black, Bcl-2 inhibitors in red, MDM2 inhibitors in cyan, LFA inhibitors in green, and Xiap inhibitors in blue. Variable or descriptor map (bottom panel). For both panels the two first axis of the PCA are represented covering more than 60% of the total variance.

iPPI versus ADME/tox properties

One very important aspect when developing drug candidates is the control of some key properties such as the molecular weight and the logP to avoid undesired behavior like promiscuous binding, poor solubility, or toxicity[26]. This becomes particularly true when dealing with compounds aiming at being administrated orally. To this end, the work of Lipinski and the so-called Rule of 5 (RO5) is well known [27]. The RO5 states that drugs that are designed for oral administration should have a molecular weight below 500 Da, a logP below 5, a number of Hydrogen bond acceptors below 10, and a number of Hydrogen bond donors below 5. Although the drug design community does not always fully agree with this guideline and physico-chemical thresholds, it is however widely used. Previous studies such as the one mentioned above, carried out by Pagliaro, found that among 19 iPPI identified from the literature only 8 could survive the RO5. Although it is not clear from their paper whether they tolerated two, one or no violation to the rule, it is interesting to apply the RO5 on this bigger representative sample of 115 iPPI and compare the results with enzyme inhibitors. The figure below **(Figure 3)** shows that if one applies the RO5 with no violation tolerated, more than two third of the iPPI are rejected (69%). This would evidently make High Throughput Screening campaigns less inclined to identify new compounds capable of modulating PPI targets. If one violation is tolerated, which is most of the time the case when using the RO5, the situation is clearly not as dramatic with this time almost two third (65%) of the iPPI being accepted although it is not as high as for enzyme inhibitors (84%). In the case of two violations tolerated the situation for iPPI (91% accepted) is similar to enzyme inhibitors (99% accepted).

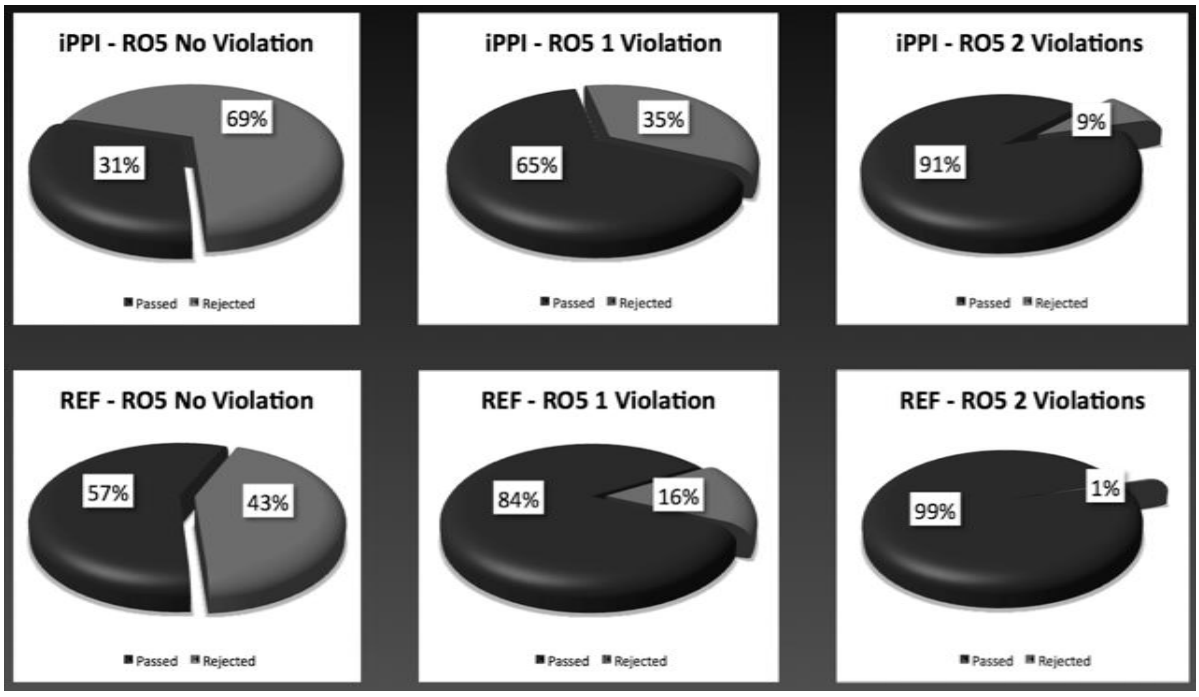
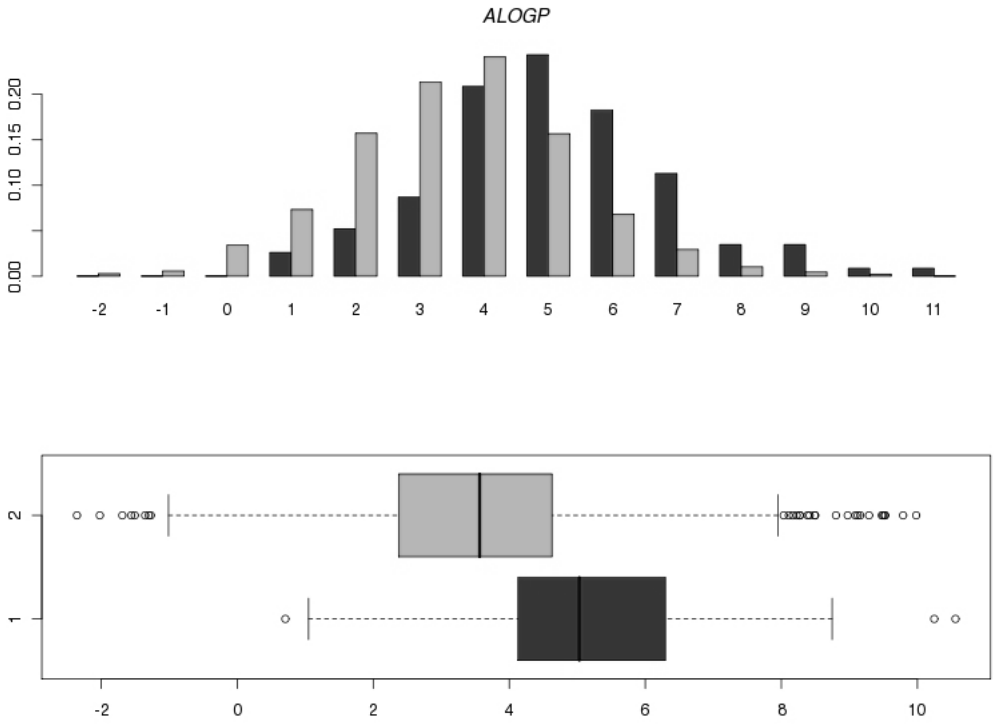
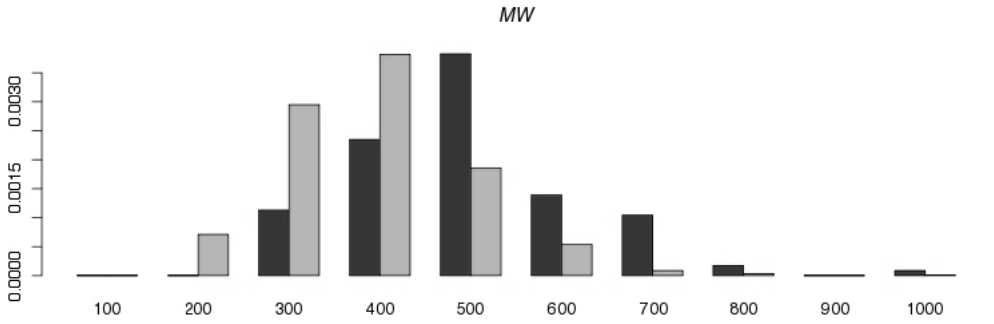
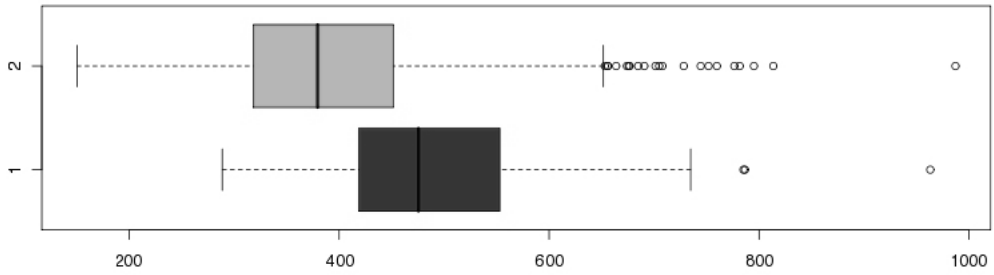


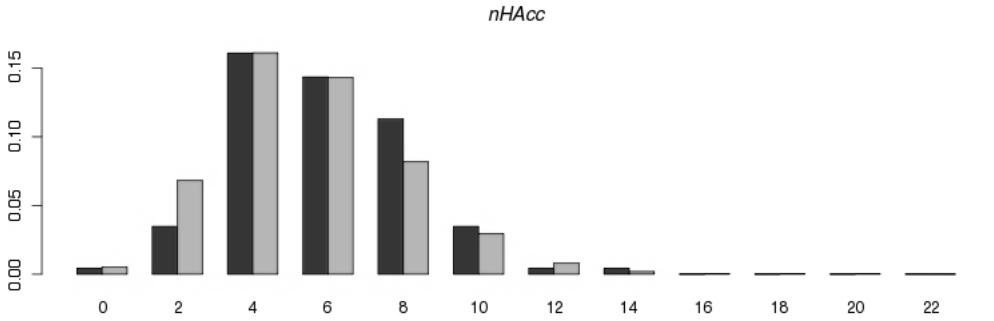
Figure 3: RO5 rule versus iPPI and enzyme inhibitors

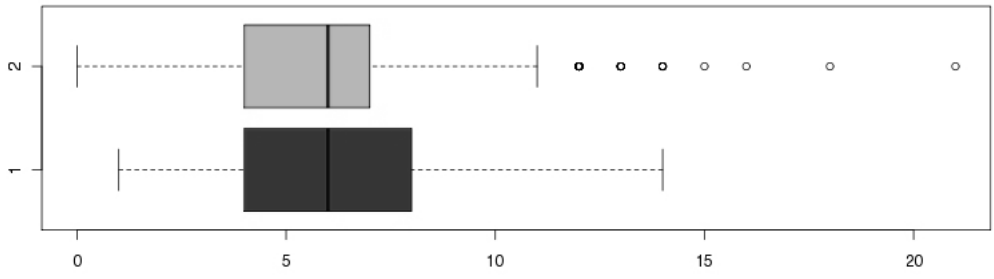
We then took into consideration the individual physico-chemical properties concerned by the RO5 and also by the Veber rule[28] (TPSA < 140 Å^2 and Number of Rotatable Bonds (RBN) < 10) using univariate distributions combining histograms and box plots **(Figure 4).** Those plots can show the most discriminative properties between the enzyme inhibitors and iPPI. As seen on the figure below, the mean value for AlogP on iPPI is 5.15, while it is 3.5 for enzyme inhibitors with a statistically significant difference (p-value << 0.05). The same is true with the molecular weight with a mean value of 496 Da for iPPI and of 388 Da for enzyme inhibitors with again a statistically significant difference (p-value << 0.05). But the other properties are either similar (nHDon, nHAcc, TPSA) or not statistically very different (RBN).

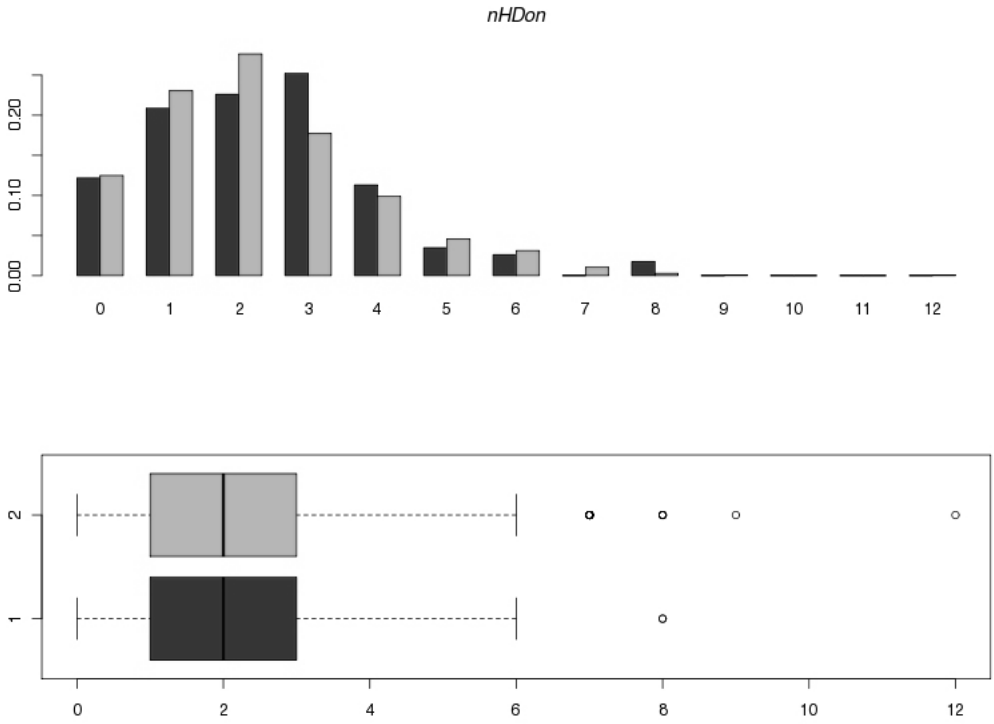



MW
0.0000
0.0015
0.0030
100
200
300
400
500
600
700
800
900
1000


1
2
200
400
600
800
1000


nHAcc
0.00
0.05
0.10
0.15
0
2
4
6
8
10
12
14
16
18
20
22


1
2
0
5
10
15
20


nHDon
0.00
0.10
0.20
0 1 2 3 4 5 6 7 8 9 10 11 12
1
2
0 2 4 6 8 10 12

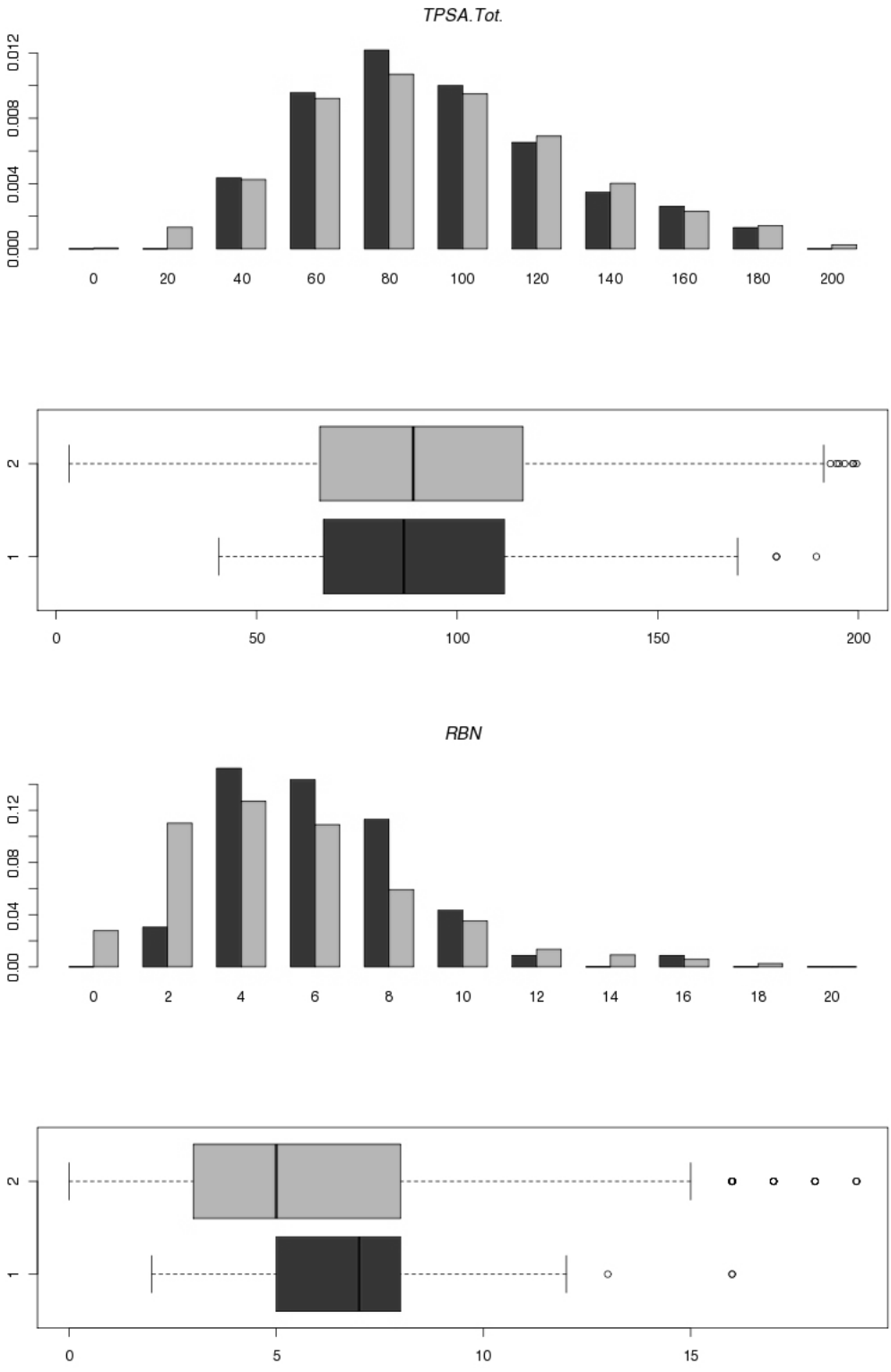
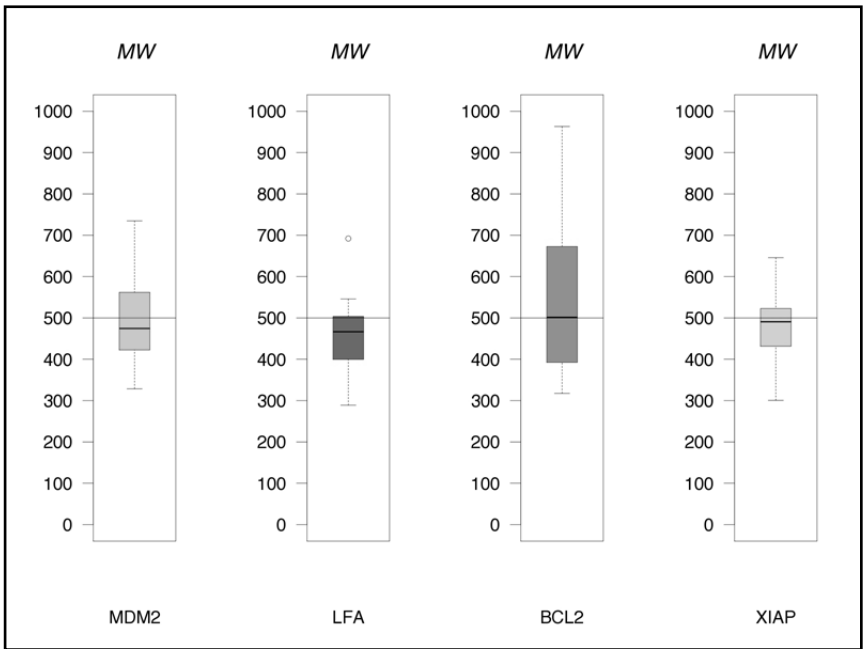
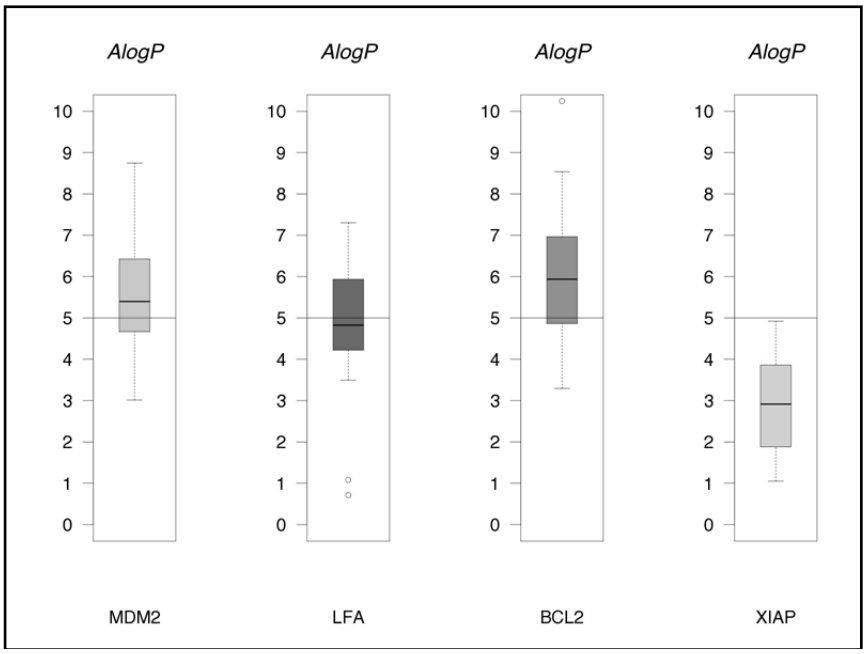


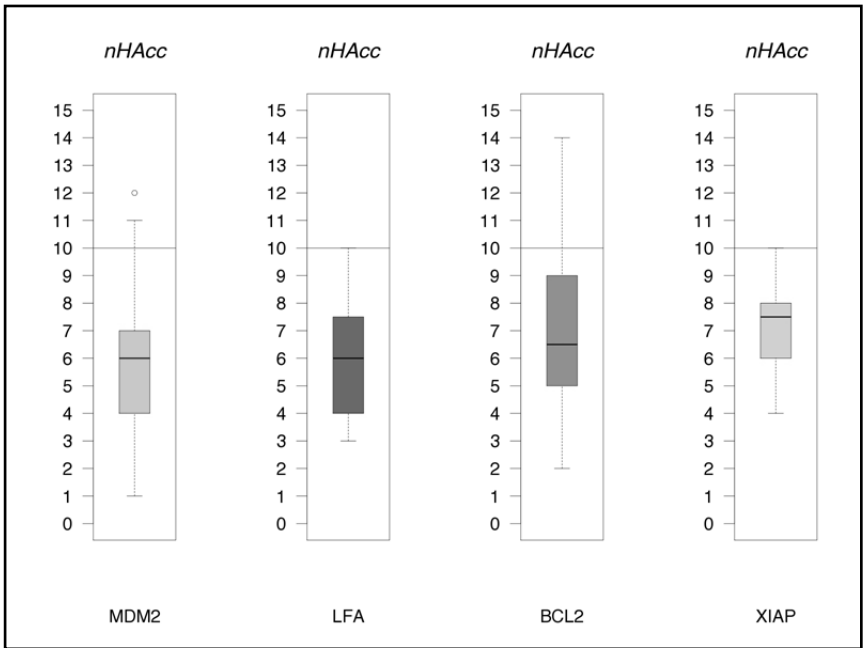
Figure 4: Double histograms and box plots for the RO5 and Veber descriptors. The enzyme inhibitors distributions are shown in green, and the iPPI are shown in red.

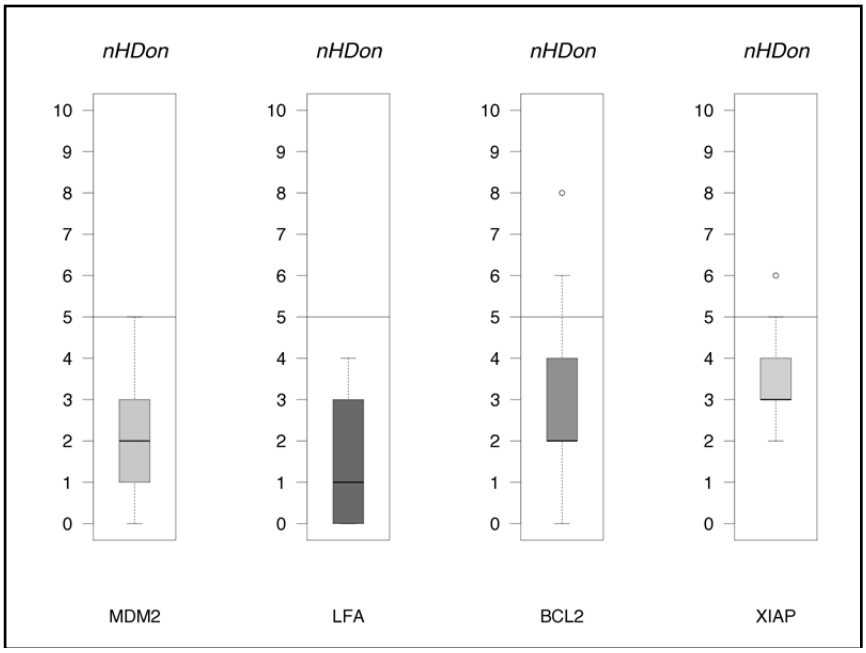
This describes a very important aspect of the propensity of iPPI for NOT being orally bioavailable, at least in the scope of the Lipinski-derived estimation of oral routing. This means that the main reason for not respecting the RO5 and therefore for having a poor propensity for being orally bioavailable is only due to too high molecular weight and to

too high logP and apparently not to the other important physico-chemical properties involved in the RO5 and Veber rules.
This is confirmed by the figure below **(Figure 5)**, which shows boxplot figures of those RO5 and Veber descriptors per PPI target along with their threshold as horizontal black lines. One can see for every PPI target, iPPI have a mean value for the molecular weight centered on about 500 Da, while there seems to be a very specific profile for Xiap compounds that are clearly not as hydrophobic as confirmed by their reduced aromatic character (mentioned above in the PCA individual map). For the other descriptors, there is clearly no problem in terms of RO5 or Veber rule violations even for some of the large compounds of the Bcl-2 family.






nHAcc
nHAcc
nHAcc
nHAcc
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
0
MDM2
LFA
BCL2
XIAP


nHDon
nHDon
nHDon
nHDon
10
9
8
7
6
5
4
3
2
1
0
MDM2
LFA
BCL2
XIAP

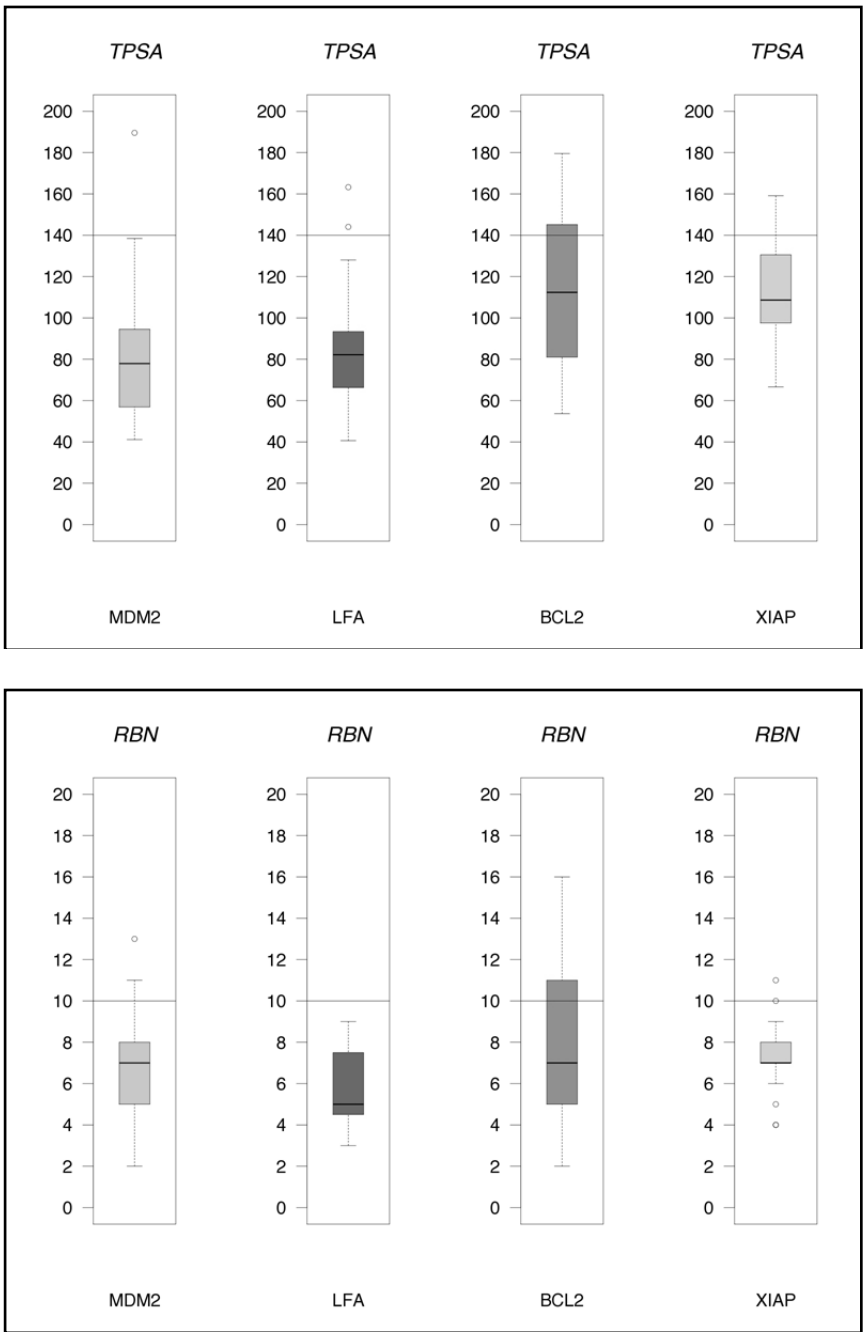


Figure 5: Individual vertical boxplot for iPPI on the 4 PPI targets using the RO5 and Veber descriptors. On each vertical boxplot is represented the corresponding threshold as horizontal black line, e.g 500 Da for MW in the RO5.

In the biplot view below **(Figure 6)** which combines molecular weight and AlogP values for iPPI per PPI target one can see that most of the iPPI do not respect one of those two descriptors and some none of them. This also means that because more than two third of iPPI respect the RO5 with one violation there must be some balance mechanism that do exist preventing most of iPPI from being concomitantly over the molecular weight and logP thresholds. The MW-AlogP biplot also shows that not all PPI targets are equally represented by their iPPI. Indeed, the Xiap compounds (blue) are all below the AlogP threshold of 5 with most of the time a moderate molecular weight.

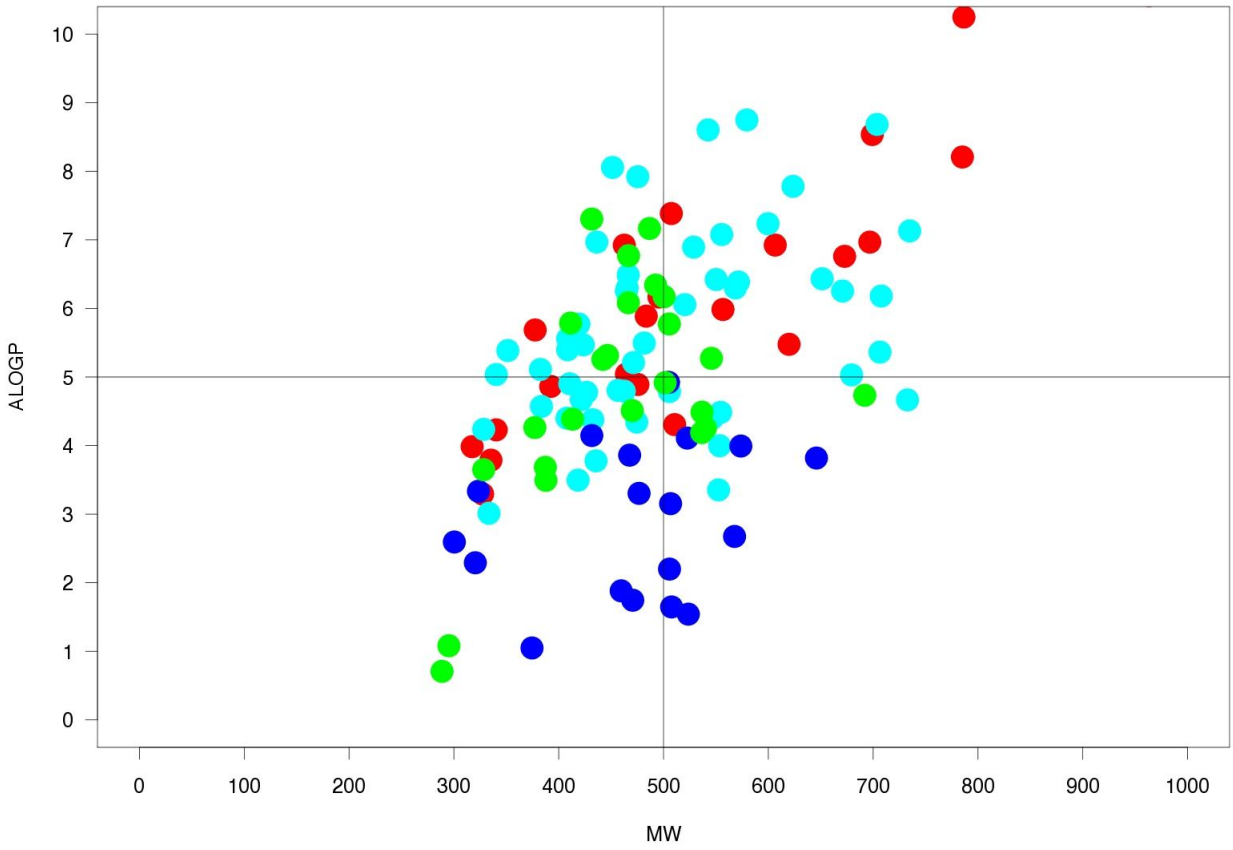


Figure 6: MW-AlogP biplot for the iPPI on the 4 PPI targets. Bcl-2 inhibitors in red, MDM2 inhibitors in cyan, LFA inhibitors in green, and Xiap inhibitors in blue.

A similar biplot figure **(Figure 7)** combining AlogP and TPSA allows one to consider the 3-75 rule that is usually applied to evaluate compound toxicity[29]. This rule stipulates that compounds with a logP below 3 and a TPSA or PSA above 75 Å^2 have better propensity for not being toxic. As can be seen on this biplot very few iPPI are in good agreement with this rule, and most of them are Xiap compounds (blue). Two of them are LFA/ICAM inhibitors (green). But this studies also shows that the most problematic compounds in terms of toxicity are in the top-left corner of the biplot (logP > 3 and TPSA < 75 Å^2), and that compounds in the top right corner (logP > 3 and TPSA > 75 Å^2) have almost the same toxicity profile than those of the bottom right corner (logP < 3 and TPSA > 75 Å^2) at least when considering all drugs regardless of their human serum albumin binding profile. Such that in the case of iPPI a fair amount of compounds is not associated to any toxicity profiles at least in the scope of this dataset.

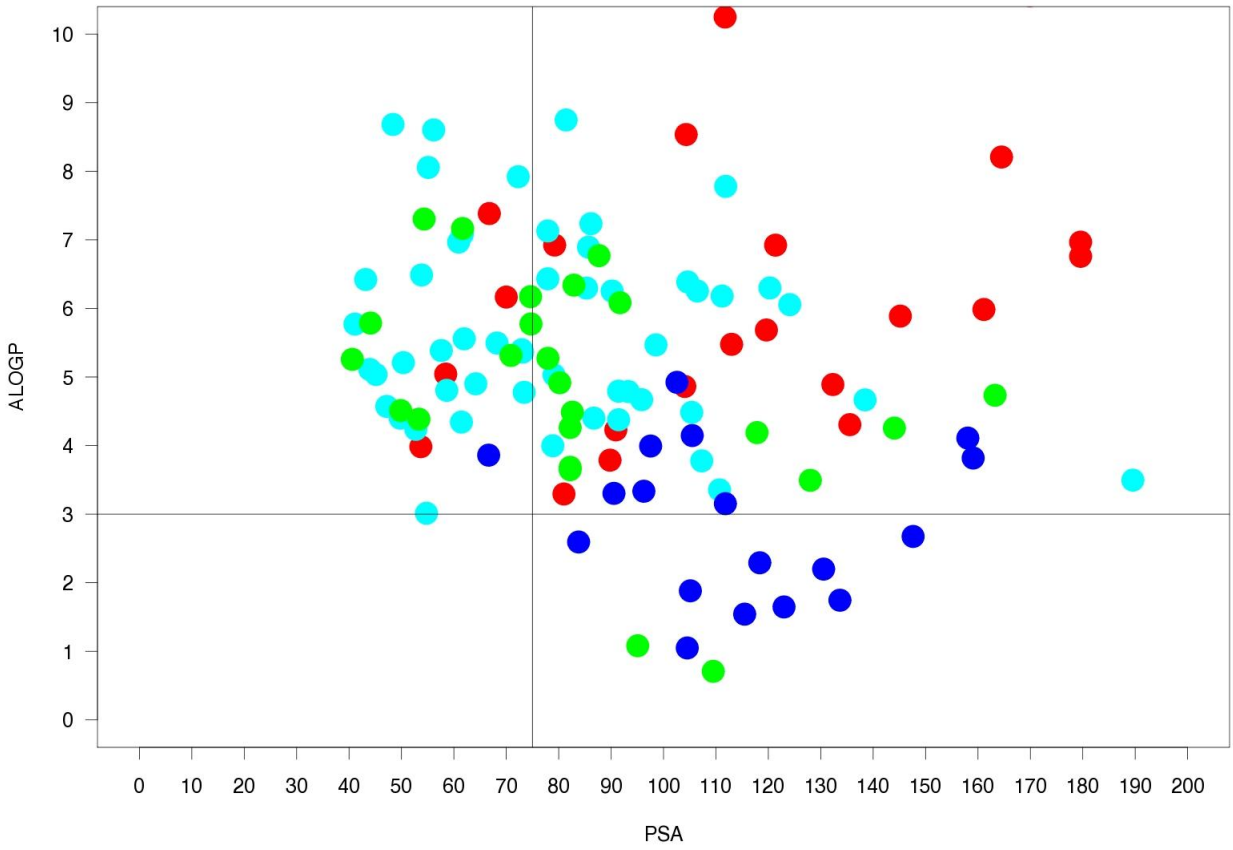
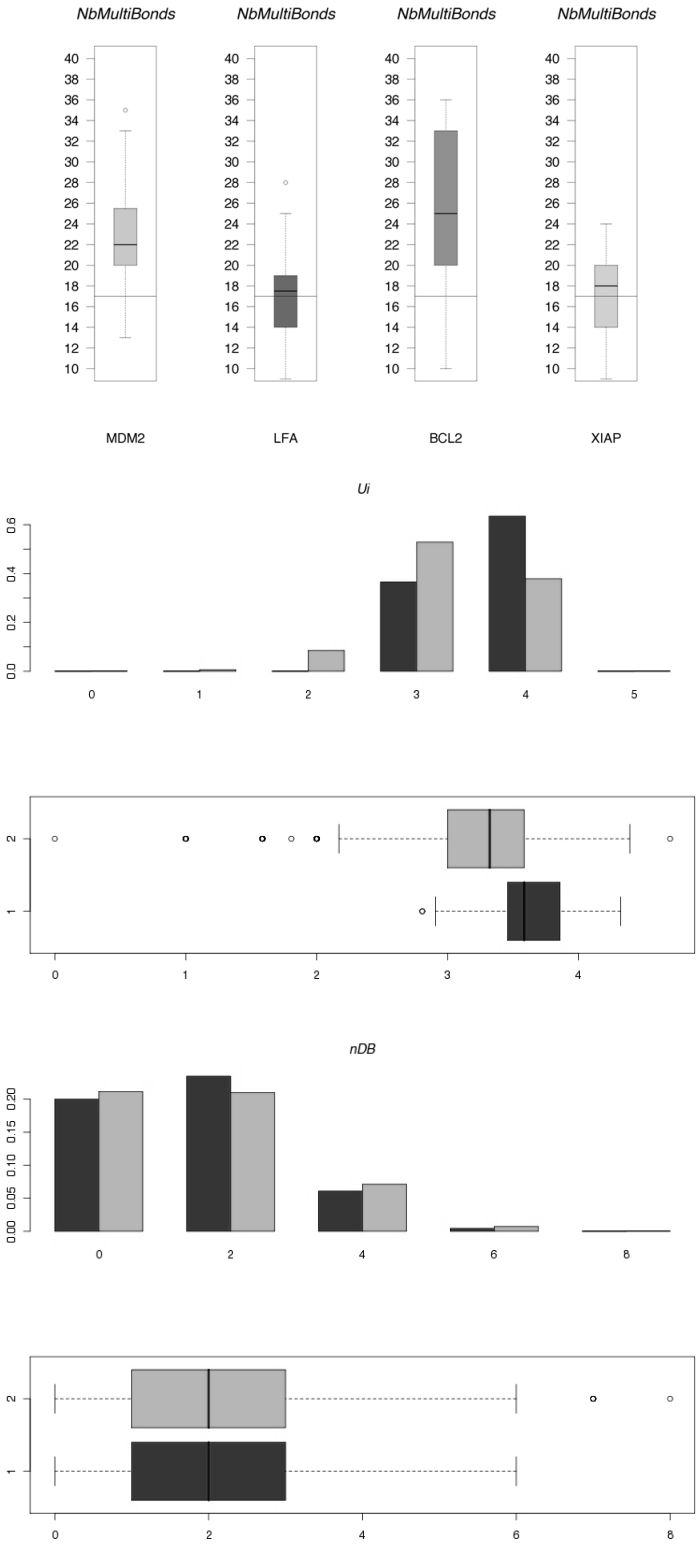


Figure 7: 3-75 Rule for the iPPI on the 4 PPI targets. Bcl-2 inhibitors in red, MDM2 inhibitors in cyan, LFA inhibitors in green, and Xiap inhibitors in blue.

iPPI versus aromaticity

As mentioned in the introduction, various studies have described the importance of aromaticity when it comes to designing iPPI, phenyl moieties are over represented, fragment such as biphenyl are known for years to be a good start to some extent when initiating a chemistry project on PPI targets. The best example is the identification of ABT-737 on Bcl-2 using fragment-based approach with, among the first chemical fragment probe, a biphenyl group that was latter on used to design the full compound[30]. In our previous work we demonstrated the importance of multiple bonds for iPPI through the identification of the unsaturation index descriptor (Ui) to which, most of all, the number of aromatic bonds is contributing. Here, we have used a combination of several univariate distributions to confirm, on this new and more populated dataset, the importance of the Ui descriptor to discriminate between iPPI and none iPPI compounds (here enzyme inhibitors) and therefore the importance of aromatic bonds. As can be seen on figure 8, Ui and the number of aromatic bonds (nAB) have statistically significant different mean values between iPPI and enzyme inhibitors, while the number of double bonds (nDB), which also contributes to Ui, has not. The figure also shows that the significant higher number of rings (nCIC) among iPPI is certainly due to a higher number of benzene-like rings (nBnz) confirming the prevalence of phenyl groups among iPPI.

The vertical boxplot panel for Ui describing PPI target individually shows that the number of multiple bonds is particularly high for MDM2 and Bcl-2 compounds, while it is moderate for LFA and Xiap compounds. Aromaticity ~~and therefore hydrophobicity~~ seems to be a price to pay to reach a sufficient level of potency for the two former targets.



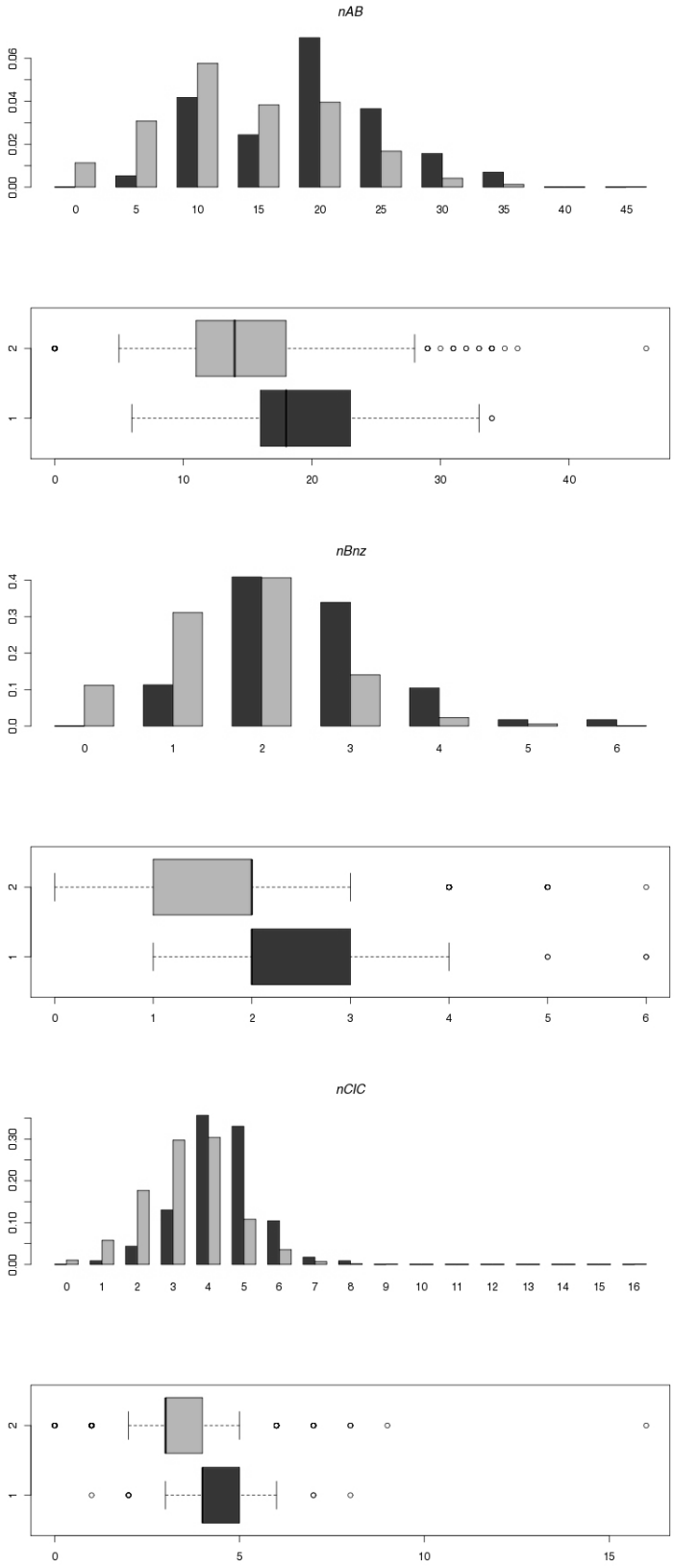


Figure 8: Double histograms and boxplot for the enzyme inhibitors (green) and iPPI (red) using the bond valence and rings descriptors.

iPPI versus chemical complexity

A characteristic that is now commonly inspected within chemical compounds is their level of structural complexity. Various ways exist to evaluate this complexity; here we have considered two of them, the ratio of sp3 carbons and the number of stereo centers. It is clear from the figure below **(Figure 9)**, that iPPI active on target Xiap/smac have a specific profile with respect to other PPI targets and to enzymes. It is non-only characterized by a significantly higher sp3 carbon ratio but also by a higher number of chiral centers. This is evidently in total agreement with the less pronounced aromatic character of the iPPI on Xiap compared to iPPI on Bcl-2 and MDM2 (as highlight within the principal component analysis paragraph above), which are characterized by a higher proportion of sp2 atoms. This, combined to a significant higher ratio of heteroatoms (data not shown) could explain the lower logP found for iPPI on Xiap.

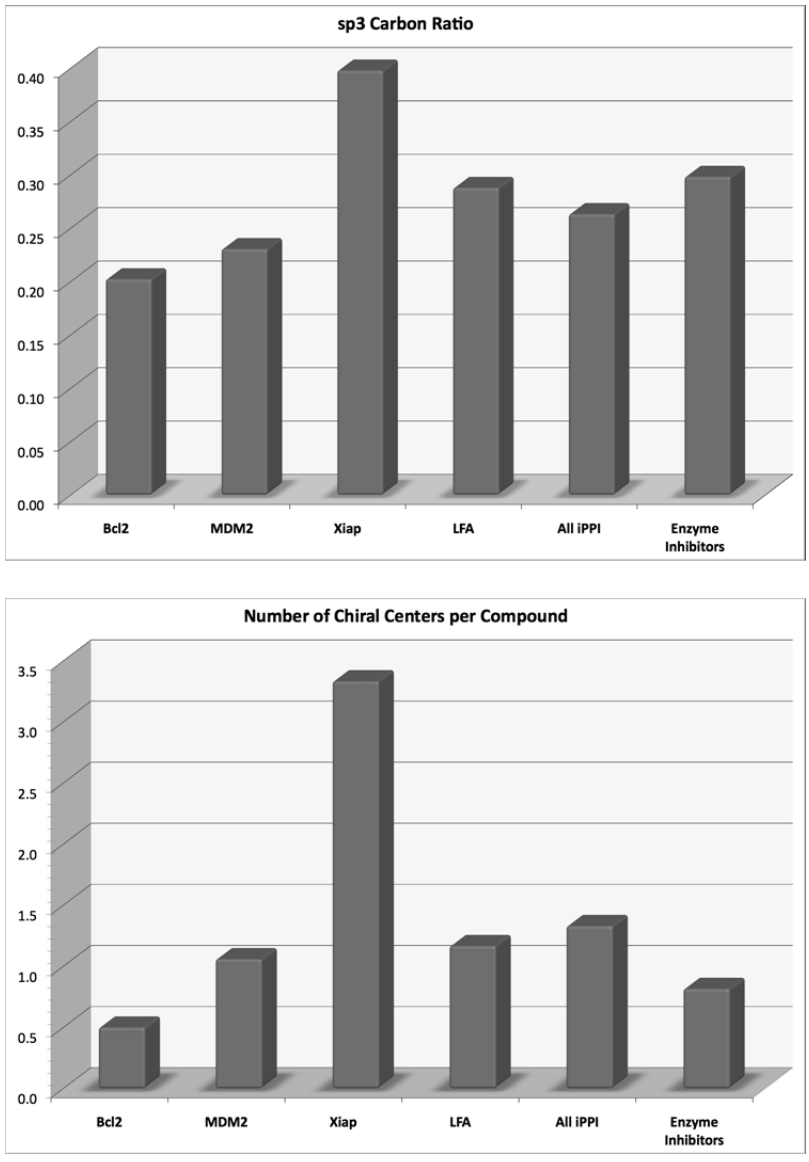


Figure 9: Structure complexity of iPPI versus enzyme inhibitors. The figure shows the sp3 carbon ratio (top panel) and the number chiral centers (bottom panel).

iPPI versus molecular shape

Molecular shape is known to be key for binding regardless of the targets and help to maximize the surface complementarities between ligand and protein. Specific ligand shapes have been proposed to be appropriate in the binding of a variety of proteins [31, 32]. In the field of PPI, Neugebauer et al. identified the SHP2 descriptor as important for iPPI potential. Yet, this descriptor is known for years as a shape descriptor. Our group has also shown the importance of molecular shape for iPPI but quantitatively. In our previous work we identified the molecular shape descriptor RDF070m (Radial Distribution Function descriptor) as specific to iPPI. On this new dataset, univariate distributions **(Figure 10)** confirm the discriminative character of RDF070m toward iPPI and highlight the more important prevalence of specific shape for the MDM2 and Bcl-2 targets than for the LFA and Xiap targets due to different distributions. It seems that ramified structures and therefore more radially distributed atoms such as those found for the MDM2 compounds are favored to address simultaneously the different subpockets of MDM2. This is confirmed by the work of Fuller et al[3] which showed that PPI pockets, as opposed to regular target pockets, have several sub-pockets rather than one large well defined and mostly continuous pocket.

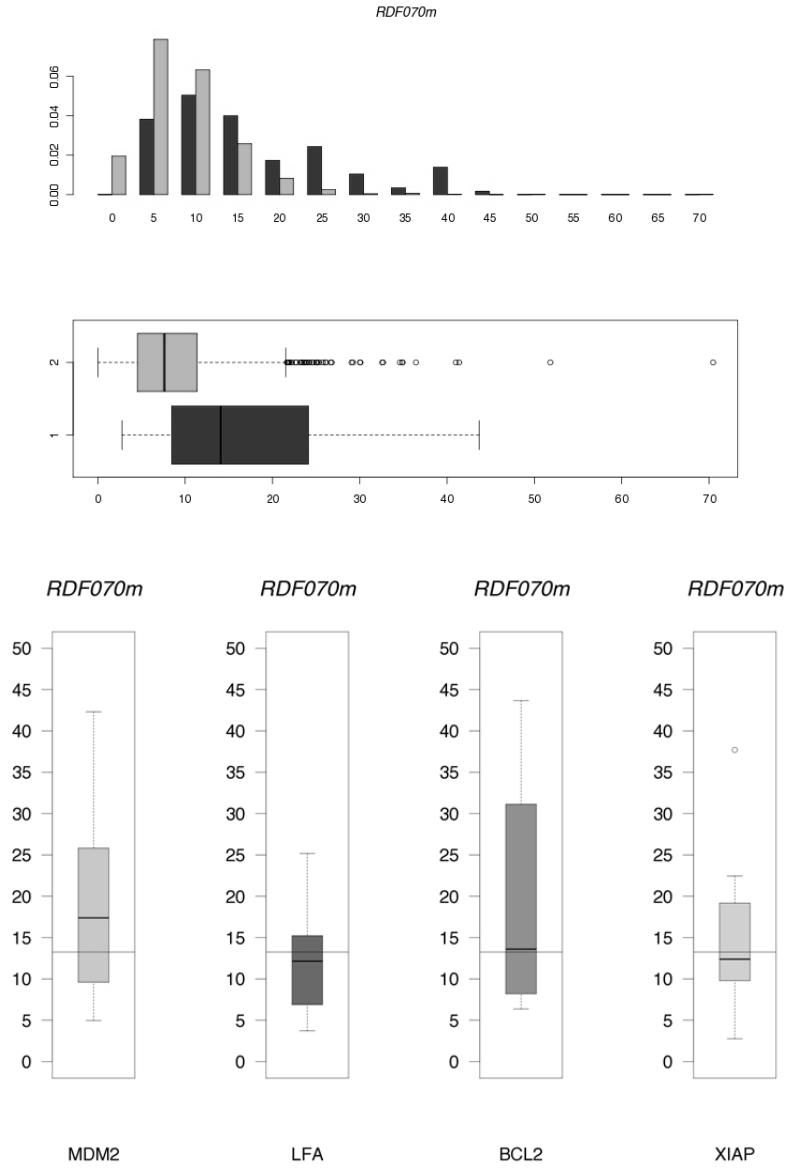
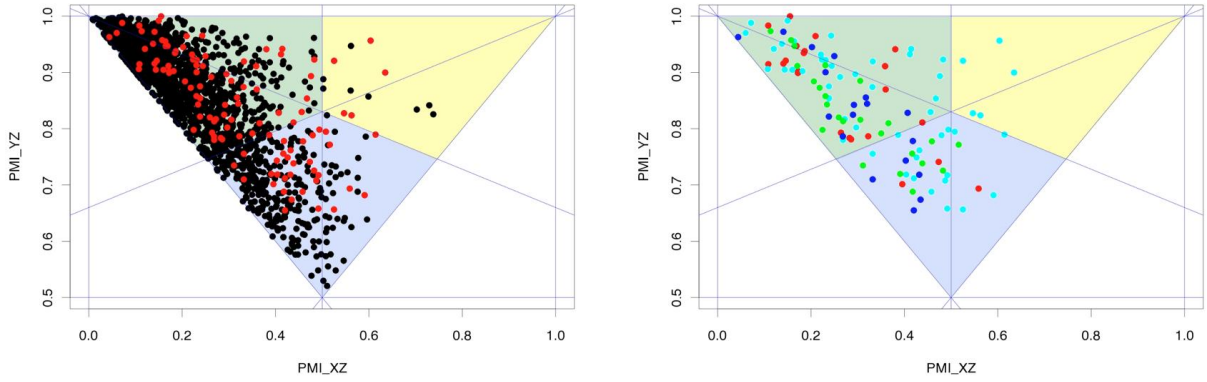


Figure 10: Double histogram and boxplot for enzyme inhibitors (green) and iPPI (red) using the molecular shape descriptor RDF070m (top panel). Vertical boxplot (bottom panel) per PPI target showing the threshold described in a previous study (13.15).

Principal moments of inertia can also be used to evaluate the molecular shape of compounds. We used the combinations of the three principal moments of inertia to evaluate the global shape of the compounds in our two datasets iPPI + enzyme inhibitors. The three principal moments of inertia were calculated using PipelinePilot 7.5, and were normalized such as to respectively divide the X and Y components by the Z component which had for consequence this triangular plot with each summit corresponding to an extreme shape. We then plotted them such as to distinguish rod- (top left corner), pancake- (bottom center), and sphere-like (top right corner) compounds as shown in the figure below **(Figure 11)**[33]. Moreover, in order to determine rod-, pancake- and sphere-like zones, we determined the equations of each summit's bisection. Using those equations we were in the position to determine a simple counting of the compounds for each zone having an inclination for the 3 extreme types of shape. In the top left panel, one can see both the enzyme inhibitors (black) and iPPI (red). It is clear from this panel that most of the enzyme inhibitors tend to occupy the

rod-like region. This is confirmed by the stacked histograms from the bottom panel of Figure 11, which shows the distribution of the compound shapes per target type, among the 3 extreme shapes. Indeed, 81% of enzyme inhibitors have a rod-like shape for globally only 67% among iPPI. On the top right panel, the same representation with only iPPI but colored by PPI target shows that there seems to be a more equal distribution between the rod-like region and the pancake-like region with much fewer compounds having a sphere-like shape inclination. But interestingly, the few examples of compounds with a sphere-like shape are all MDM2 compounds, which is confirmed by the higher average value of RDF070m for this PPI target. Interestingly, the bottom panel shows that LFA and Xiap inhibitors have very similar profiles in terms of molecular shape distributions while most of the Bcl-2 inhibitors have a rod-like shape (86%) even greater in proportion than enzyme inhibitor. One might argue it is due to the topology of the Bcl-2 binding pocket, deep and extended, which receives a α-helix from the BH3 domain of its partners that is about 20-amino acid long.



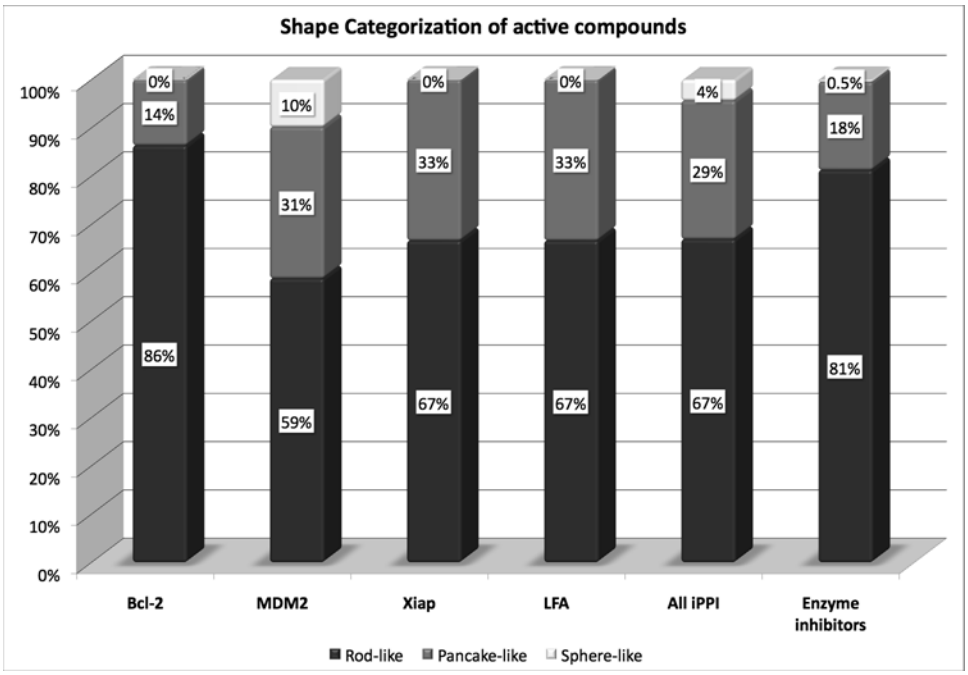


Figure 11: Principal Moments of Inertia represented as a 2D plot. The triangle plots represent the global shape of the compounds sphere-like (yellow zone), pancake-like (blue-zone), and rod-like shape (green zone). Left panel shows both enzyme inhibitor (black) and iPPI (red). Right panel shows iPPI colored by PPI targets, Bcl-2 inhibitors in red, MDM2 inhibitors in cyan, LFA inhibitors in green, and Xiap inhibitors in blue.

iPPI versus Potency

We have then analyzed the potency of those 115 iPPI versus enzyme compounds using activity bins of pIC50 as in the figure below **(Figure 12)**. Cumulated proportions (left panel) show that only 6% of iPPI have pIC50 above 8 and 26% above 7, while 16% of enzyme inhibitors have pIC50 above 8 and 37% above 7. If the RO5 is applied to the datasets with one tolerated-violation, a closer look at the compounds that passed the RO5 (central panel) highlights identical proportions of compounds among the different activity bins for the enzyme inhibitors (16%, 21%, 25%, 38%). The same is not true with iPPI for which the pIC50-above-8 bin drops to 4% and the cumulated pIC50-above-7 bin drops to 19%. But the most striking results are obtained when considering the subpopulations that did not pass the RO5 (right panel). In that configuration, the proportion among bins is still similar for enzyme inhibitors, but changes dramatically for iPPI with a cumulated proportion of 46% of iPPI with a pIC50 above 7. This means that most of the most active compounds do not respect the RO5, which highlight the hydrophobic and molecular weight price that still needs to be paid in order to increase potency to the namomolar range.

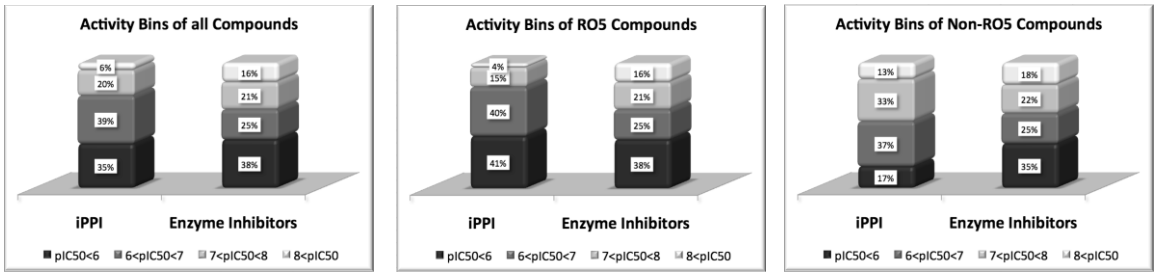


Figure 12: Normalized histograms of pIC50 bins comparing potency for enzyme inhibitors and iPPI in the context of the RO5.

An examination of the ligand efficiency (LEHA) and lipophilic efficiency (LLE) using a biplot like the figure below confirmed that observation. When considering LE (Ligand Efficiency), for which the Gibbs free energy of binding is used, one can consider that value of 0.30 kcal/mol/heavy atom is desirable to obtain oral drug candidates[34]. As LEHA is unitless and around 0.73 times the LE value [35], this brings this druglikeness threshold using LEHA value to 0.22. Conversely, average LLE values for oral drugs are in the range 5-7[26]. The biplot figure (left panel) of LEHA (x-axis) and LLE (y-axis) shows that most of iPPI (red dots) are in the bottom left corner corresponding to poorer values of LEHA and LLE, whereas a wide proportion of enzyme inhibitors (black dots) occupy more favored regions for further optimizations **(Figure 13)**. The right panel similarly shows the detail of efficiencies per PPI target. One can see that most of the iPPI respecting this efficiency profile are Xiap/smac and ICAM/LFA compounds.

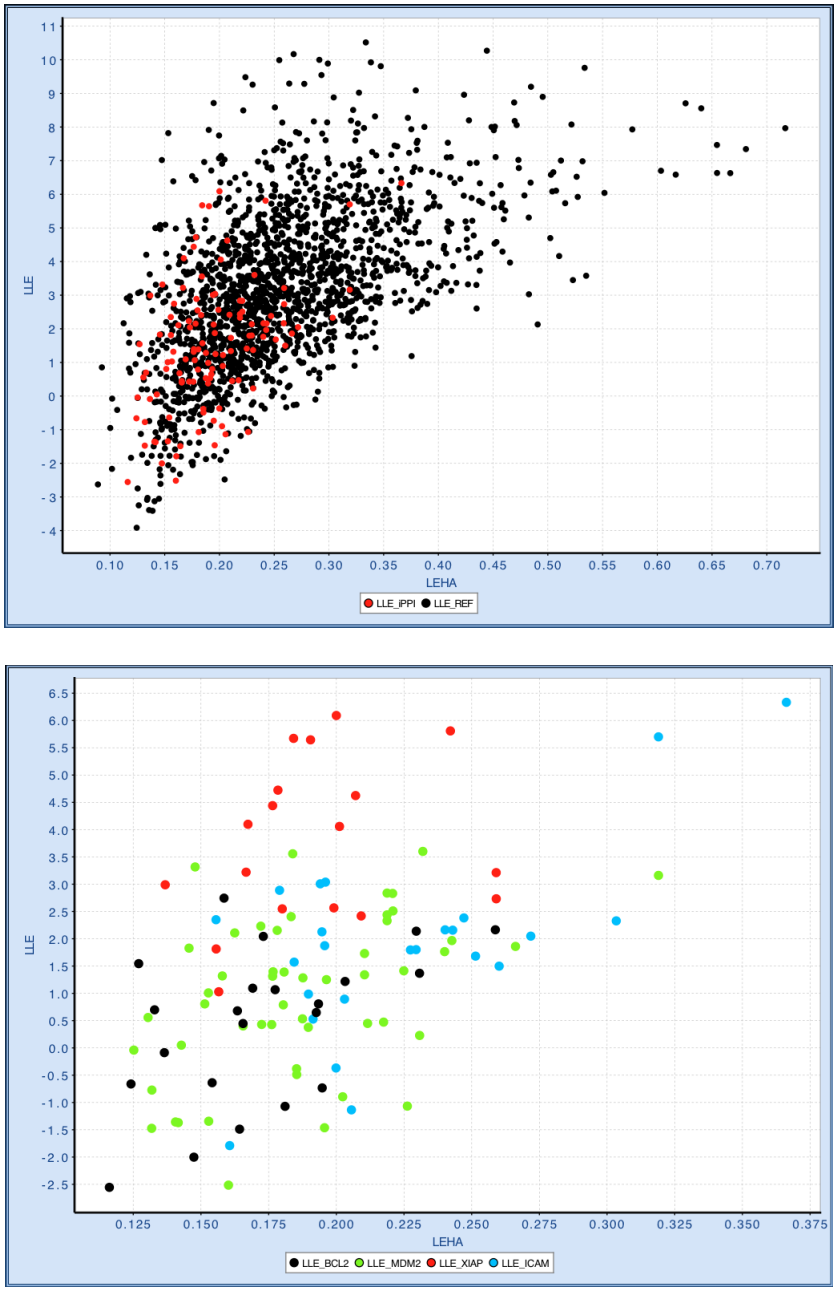


Figure 13: Ligand efficiencies (LEHA, x-axis) and lipophilic efficiencies (LLE, y-axis). Top panel shows both enzyme inhibitors (black) and iPPI (red). Bottom panel shows iPPI colored by PPI targets, Bcl-2 inhibitors in black, MDM2 inhibitors in green, LFA inhibitors in cyan, and Xiap inhibitors in red.

iPPI versus PAINS

An important aspect when dealing with active compounds during HTS binding assays is to ensure the screening is not revealing false positive compounds that could mislead chemists and biologists in the wrong direction. To this end, the work of Baell on Pan Assay INterference compounds - PAINS - is of particular importance to flag compounds containing fragments that are known to contribute to false positives in a wide range of biochemical assays[36]. Three sets of fragments were identified by Baell A, B, and C, A being the set of the most predominantly observed fragment-associated false positives. Using the program FAFDrugs2[37], we detected the PAINS corresponding to A, B, and C subsets in our two datasets. Figure 14 shows all the PAINS that are present within iPPI for the three subsets of PAINS fragments and the proportions of those PAINS for each PPI target and for enzymes. PPI inhibitors have a cumulated proportions of PAINS equal to 23% while it is only 5% for enzyme inhibitors, but the proportion of A-PAINS (fragments from subset A), which are again the most often observed PAINS, is similar (3%) for both iPPI and enzyme inhibitors. Interestingly, Xiap and LFA compounds have no PAINS at all. The most often present types of PAINS for iPPI are therefore for subsets B and C, and more specifically catechol_A for BCL-2 inhibitors (91% of B-PAINS), and anil_NH_alk_B for MDM2 inhibitors (100% of C-PAINS). So, not all PAINS from those subsets B and C are equally highlighted. The case of catechol_A is particularily interesting because it is present in a wide variety of BCL-2 inhibitors including gossypol derivatives among which some are in clinical trial in human (e.g AT101). So, information about PAINS associated compounds must be known but should be interpreted with caution.

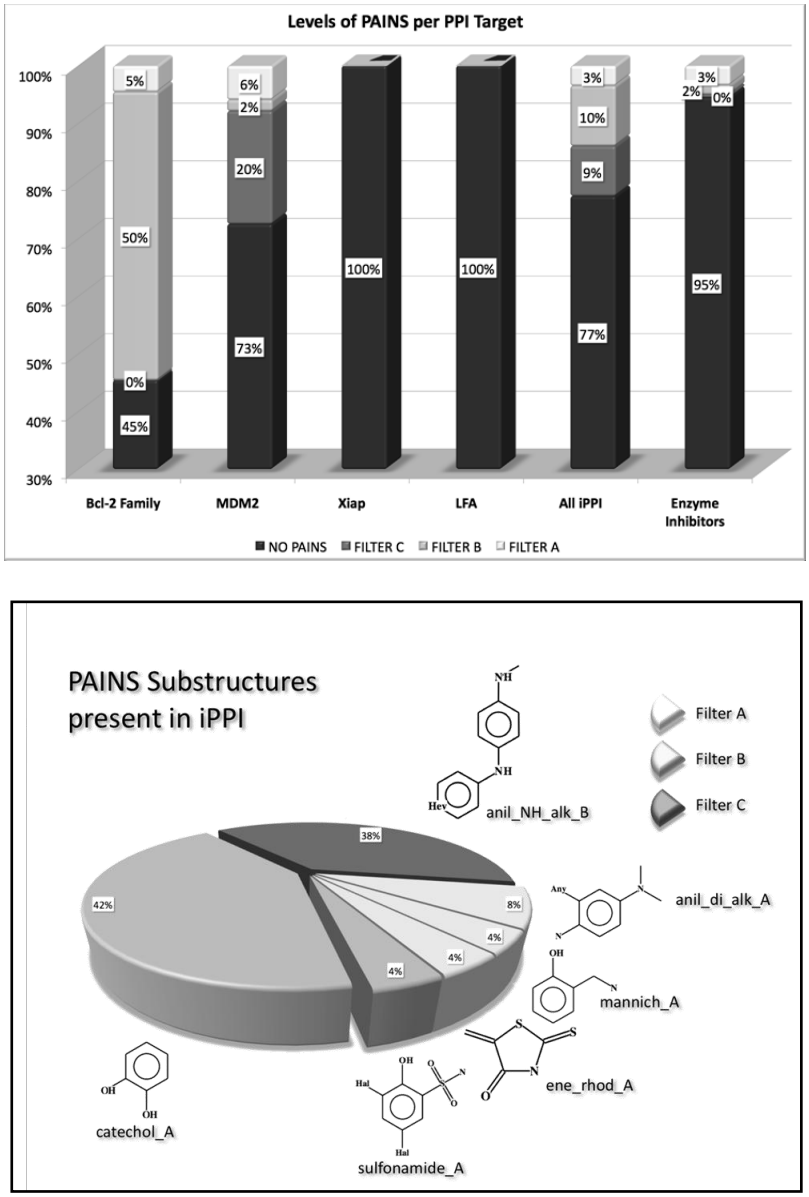


Figure 14: Levels of PAINS on the datasets. (Top panel) Proportions of compounds flagged as PAINS for each target type and for the three types of PAINS; A, B, and C. (Bottom panel) Pie chart showing the PAINS fragments among iPPI.

Conclusion

Designing or identifying iPPI is known to be a difficult task. ADME considerations will also have to be taken into account at some point in the development if one wishes to go beyond chemical biology projects or proof of concepts. The examination of success stories can be used to rationalize the properties of existing iPPI and learn how to maximize the chances of success. Closer look at existing examples are necessary because they demonstrate major trends in the design and some balance that need to be found to reduce failure. The present case study analysis demonstrates that we have a long way to go in order to propose more drug/lead-like compound as iPPI because the major tendency is still to increase the hydrophobicity, the aromaticity, and the molecular weight of these compounds to maximize potency, but to levels that might not permit

further optimization (by contrast to new trends in enzyme inhibitor design). Indeed, already at the hit stage, iPPIs display properties that suggest higher risk of promiscuous binding or even toxicity. Yet, among the presented examples, some compounds have acceptable properties and some targets e.g Xiap seem to be exempted from the hydrophobic and weight prices to pay for potency and efficacy. Lessons must be learnt from those examples both for iPPI drug-likeness and protein druggability or ligandability, which is the capacity of a protein to bind a small compound, in order to develop novel drug candidates for this new class of challenging targets. This opens the way of good practice both in terms of iPPI design but also in terms of PPI target categorization and the determination what type of PPI target must be prioritized[38]. To this end, a major leap must be taken toward a better comprehension of the different PPI categories. Must this categorization be pocket-driven, binder-driven (α-helix-, β-turn-, or β-sheet-mediated PPI), or compound-driven? And most of all to which category correspond both PPI ligandability and PPI druggability? Most likely, this learning phase will allow the scientific community to explore and to design more ADMET-friendly compound collections dedicated to the direct inhibition of protein-protein interactions within the next few years.

References

[1] Venkatesan K, Rual JF, Vazquez A, Stelzl U, Lemmens I, Hirozane-Kishikawa T, Hao T, Zenkner M, Xin X, Goh KI, Yildirim MA, Simonis N, Heinzmann K, Gebreab F, Sahalie JM, Cevik S, Simon C, de Smet AS, Dann E, Smolyar A, Vinayagam A, Yu H, Szeto D, Borick H, Dricot A, Klitgord N, Murray RR, Lin C, Lalowski M, Timm J, Rau K, Boone C, Braun P, Cusick ME, Roth FP, Hill DE, Tavernier J, Wanker EE, Barabasi AL, Vidal M. An empirical framework for binary interactome mapping. Nat Methods, 2009; 6: 83-90.

[2] Stumpf MPH, Thorne T, de Silva E, Stewart R, An HJ, Lappe M, Wiuf C. Estimating the size of the human interactome. Proceedings of the National Academy of Sciences, 2008; 105: 6959-6964.

[3] Fuller JC, Burgoyne NJ, Jackson RM. Predicting druggable binding sites at the protein-protein interface. Drug discovery today, 2009; 14: 155-161.

[4] Perkins JR, Diboun I, Dessailly BH, Lees JG, Orengo C. Transient protein-protein interactions: structural, functional, and network properties. Structure (London, England : 1993), 2010; 18: 1233-1243.

[5] Nooren IMA, Thornton JM. Diversity of protein-protein interactions. The EMBO journal, 2003; 22: 3486-3492.

[6] Jones S, Thornton JM. Principles of protein-protein interactions. Proceedings of the National Academy of Sciences of the United States of America, 1996; 93: 13-20.

[7] Clackson T, Wells JA. A hot spot of binding energy in a hormone-receptor interface. Science, 1995; 267: 383-386.

[8] Ma B, Nussinov R. Trp/Met/Phe hot spots in protein-protein interactions: potential targets in drug design. Current topics in medicinal chemistry, 2007; 7: 999-1005.

[9] Bullock BN, Jochim AL, Arora PS. Assessing helical protein interfaces for inhibitor design. J Am Chem Soc, 2011; 133: 14220-3.
[10] Lanzarotti E, Biekofsky RR, Estrin DoA, Marti MA, Turjanski AnG. Aromatic-aromatic interactions in proteins: beyond the dimer. Journal of Chemical Information and Modeling, 2011; 51: 1623-1633.
[11] Davis FP, Sali A. The overlap of small molecule and protein binding sites within families of protein structures. PLoS computational biology, 2010; 6: e1000668.
[12] Kozakov D, Hall DR, Chuang GY, Cencic R, Brenke R, Grove LE, Beglov D, Pelletier J, Whitty A, Vajda S. Structural conservation of druggable hot spots in protein-protein interfaces. Proc Natl Acad Sci U S A, 2011; 108: 13528-33.
[13] Berg T. Small-molecule inhibitors of protein-protein interactions. Curr Opin Drug Discov Devel, 2008; 11: 666-74.
[14] Fry DC. Drug-like inhibitors of protein-protein interactions: a structural examination of effective protein mimicry. Curr Protein Pept Sci, 2008; 9: 240-7.
[15] Fry DC. Protein-protein interactions as targets for small molecule drug discovery. Biopolymers, 2006; 84: 535-552.
[16] Wilson AJ. Inhibition of protein-protein interactions using designed molecules. Chemical Society reviews, 2009; 38: 3289-3300.
[17] Wilson CGM, Arkin MR, Vassilev L, Fry D eds. Small-Molecule Inhibitors of IL-2/IL-2R: Lessons Learned and Applied. Springer Berlin Heidelberg: Berlin, Heidelberg 2010.
[18] Higueruelo AP, Schreyer A, Bickerton GR, Pitt WR, Groom CR, Blundell TL. Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. Chem Biol Drug Des, 2009; 74: 457-67.
[19] Bourgeas R, Basse M-J, Morelli X, Roche P. Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. PLoS ONE, 2010; 5: e9598.
[20] Neugebauer A, Hartmann RW, Klein CD. Prediction of protein-protein interaction inhibitors by chemoinformatics and machine learning methods. Journal of Medicinal Chemistry, 2007; 50: 4665-4668.
[21] Morelli X, Bourgeas R, Roche P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). Curr Opin Chem Biol, 2011; 15: 475-81.
[22] Reynes C, Host H, Camproux AC, Laconde G, Leroux F, Mazars A, Deprez B, Fahraeus R, Villoutreix BO, Sperandio O. Designing focused chemical libraries enriched in protein-protein interaction inhibitors using machine-learning methods. PLoS Comput Biol, 2010; 6: e1000695.
[23] Sperandio O, Reynes CH, Camproux AC, Villoutreix BO. Rationalizing the chemical space of protein-protein interaction inhibitors. Drug Discov Today, 2010; 15: 220-9.
[24] Broos K, Trekels M, Jose RA, Demeulemeester J, Vandenbulcke A, Vandeputte N, Venken T, Egle B, De Borggraeve WM, Deckmyn H, De Maeyer M. Identification of a Small Molecule That Modulates Platelet Glycoprotein Ib-von Willebrand Factor Interaction. J Biol Chem, 2012; 287: 9461-72.
[25] Liu T, Lin Y, Wen X, Jorissen RN, Gilson MK. BindingDB: a web-accessible database of experimentally determined protein-ligand binding affinities. Nucleic Acids Res, 2007; 35: D198-201.
[26] Leeson PD, Springthorpe B. The influence of drug-like concepts on decision-making in medicinal chemistry. Nat Rev Drug Discov, 2007; 6: 881-90.

[27] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Advanced drug delivery reviews, 2001; 46: 3-26.
[28] Veber DF, Johnson SR, Cheng HY, Smith BR, Ward KW, Kopple KD. Molecular properties that influence the oral bioavailability of drug candidates. J Med Chem, 2002; 45: 2615-23.
[29] Hughes JD, Blagg J, Price DA, Bailey S, Decrescenzo GA, Devraj RV, Ellsworth E, Fobian YM, Gibbs ME, Gilles RW, Greene N, Huang E, Krieger-Burke T, Loesel J, Wager T, Whiteley L, Zhang Y. Physiochemical drug properties associated with in vivo toxicological outcomes. Bioorg Med Chem Lett, 2008; 18: 4872-5.
[30] Oltersdorf T, Elmore SW, Shoemaker AR, Armstrong RC, Augeri DJ, Belli BA, Bruncko M, Deckwerth TL, Dinges J, Hajduk PJ, Joseph MK, Kitada S, Korsmeyer SJ, Kunzer AR, Letai A, Li C, Mitten MJ, Nettesheim DG, Ng S, Nimmer PM, O'Connor JM, Oleksijew A, Petros AM, Reed JC, Shen W, Tahir SK, Thompson CB, Tomaselli KJ, Wang B, Wendt MD, Zhang H, Fesik SW, Rosenberg SH. An inhibitor of Bcl-2 family proteins induces regression of solid tumours. Nature, 2005; 435: 677-81.
[31] Kortagere S, Krasowski MD, Ekins S. The importance of discerning shape in molecular pharmacology. Trends in pharmacological sciences, 2009; 30: 138-147.
[32] Akritopoulou-Zanze I, Metz JT, Djuric SW. Topography-biased compound library design: the shape of things to come? Drug discovery today, 2007; 12: 948-952.
[33] Akella LB, DeCaprio D. Cheminformatics approaches to analyze diversity in compound screening libraries. Curr Opin Chem Biol, 2010; 14: 325-30.
[34] Chessari G, Woodhead AJ. From fragment to clinical candidate--a historical perspective. Drug Discov Today, 2009; 14: 668-75.
[35] Reynolds CH, Tounge BA, Bembenek SD. Ligand binding efficiency: trends, physical basis, and implications. J Med Chem, 2008; 51: 2432-8.
[36] Baell JB, Holloway GA. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. J Med Chem, 2010; 53: 2719-40.
[37] Lagorce D, Maupetit J, Baell J, Sperandio O, Tuffery P, Miteva MA, Galons H, Villoutreix BO. The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. Bioinformatics, 2011; 27: 2018-20.
[38] Surade S, Blundell TL. Structural biology and drug discovery of difficult targets: the limits of ligandability. Chem Biol, 2012; 19: 42-50.